

GRSciColl: How the Registry of Scientific Collection can help improve system interoperability and link data from different sources

Ms Marie Grosjean¹, M. Tim Robertson¹, M. Morten Høfft¹, M. Marcos Lopez Gonzalez¹, Ms. Andrea Hahn¹
¹GBIF Secretariat, , Denmark

GRSciColl (<https://www.gbif.org/grscicoll>), the Registry of Scientific Collections, is a global, comprehensive, community-curated clearinghouse of collections information originally developed by the Consortium of the Barcode of Life (CBOL) and was hosted by the Smithsonian Institution until 2019. It is now hosted and maintained as a component within the Global Biodiversity Information Facility (GBIF) registry.

One of the initial goals of GRSciColl was to improve interoperability among databases by providing codes and machine-readable identifiers for institutions and collections.

Today, GRSciColl contains more than 8,000 institutions and 6,000 collections and around 15,000 codes and 48,000 identifiers referencing these institutions and collections. Some of the identifiers available on GRSciColl refer to other systems such as ROR (<https://ror.org/>) and Index Herbariorum (<http://sweetgum.nybg.org/science/ih/>) which facilitates interoperability and cross-referencing.

In 2020, GRSciColl was enhanced with the Collection API Lookup service (<https://www.gbif.org/developer/registry#lookup>). This service enables users to find collections and institutions based on codes and identifiers. The Lookup service is now used to link specimen-related occurrence records within GBIF.org to GRSciColl entries when possible, making use of the various identifiers associated with records, such as collection identifiers and codes. This allows institutions to aggregate metrics for their digitized specimens regardless of how they were published on GBIF. There are currently more than 100 million specimen-related occurrences linked to GRSciColl entries.

GRSciColl is actively maintained and enhanced with new features. In 2022, we are focusing our efforts on improving the registry content. Longer-term, we would like to explore how best to integrate DOI assignment to GRSciColl entries and whether to expand the GBIF literature tracking system (<https://www.gbif.org/literature-tracking>) to credit institutions and collections.

This presentation will introduce some of the challenges experienced when building this global registry, with a focus on issues relating to different identifier formats used with the relevant data objects.

